PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/00, 15/63, 15/09

A1

(11) International Publication Number: **WO 00/44892**

(43) International Publication Date: 3 August 2000 (03.08.00)

(21) International Application Number: PCT/US00/01662

(22) International Filing Date: 26 January 2000 (26.01.00)

(30) Priority Data:
60/117,690 29 January 1999 (29.01.99) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/117,690 (CIP)
Filed on 26 January 1999 (26.01.99)

(71) Applicant *(for all designated States except US)*: TRANXENOGEN, INC. [US/US]; P.O. Box 707 MO, Shrewsbury, MA 01545 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: DITULLIO, Paul, A. [US/US]; 42 Juniper Brook Road, Northboro, MA 01532 (US).

(74) Agent: BEATTIE, Ingrid, A.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States: AU, CA, NZ, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CONTROLLED EXPRESSION OF HETEROLOGOUS PROTEINS IN THE MAMMARY GLAND OF A TRANSGENIC ANIMAL

(57) Abstract

The invention features an isolated nucleic acid containing a promoter region derived from the human lactoferin gene operably linked to a heterologous sequence, methods of expressing transgenes, and transgenic animals containing lactoferin–derived promoter regions.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Controlled Expression of Heterologous Proteins in the Mammary Gland of a Transgenic Animal

TECHNICAL FIELD

This invention relates to expression of gene expression in mammary gland tissue.

BACKGROUND

This application claims priority to U.S. Provisional Patent Application No. 60/117,690.

The field of transgenics has grown rapidly since the initial experiments describing the introduction of foreign DNA into the developing zygote or embryo (Brinster, R.L. et al., Proc. Natl. Acad. Sci. USA 82:4438-4442 (1985); Wagner et al., U.S. 4,873,191 (1989)). Transgenic technology has been applied to both laboratory and domestic species for the study of human diseases (Synder, B.W., et al., Mol. Reprod. and Develop. 40:419-428 (1995)), production of pharmaceuticals in milk (Ebert, K.M. and J.P. Selgrath, "Changes in Domestic Livestock through Genetic Engineering" in Applications in Mammalian Development, Cold Spring Harbor Laboratory Press, 1991), to develop improved agricultural stock (see, for example, Ebert, K.M. et al., Animal Biotechnology 1:145-159 (1990)) and xenotransplantation (Osman, N., et al., Proc. Natl. Acad. Sci USA 94:14677-14682 (1997)). A crucial step in the development of transgenic animals is the construction of the vector or cassette to be microinjected. The ultimate utility or value of the transgenic animal is dependent on the specificity and strength of the promoter being used to express the gene of interest. This fact is particularly evident in utilizing the mammary gland of transgenic animals for the production of pharmaceuticals.

Researchers aiming to produce pharmaceuticals in the milk of lactating transgenic animals focused on the cloning and characterization of the genes associated with the major milk proteins from the domestic species and common laboratory animals. For example, the genes for goat beta casein (Roberts, B. et al., Gene 121:255-262 (1992)) and sheep beta lactoglobulin (Simons, J.P. et al., Nature 328:530-532 (1987)) were isolated and used to produce transgenic mice to demonstrate the ability to direct expression to the mammary gland. In both cases, the protein product was detected in the milk, however, the expression was highly variable and not completely limited to the mammary gland. These

experiments clearly demonstrated that crucial control elements were not present in the vectors to correctly direct expression of the gene. This was further illustrated when a heterologous protein coding sequence was attached to a milk specific promoter (Wright, G., et al. Biotechnology 9:830-834 (1991); Ebert K.M. et al., Biotechnology 9:835-838 (1991)). In addition to the problem of inconsistent or non-tissue specific expression, researchers found that some transgenic animals over-expressed the target protein which caused problems with milk production (Shamay, A., et al., Transgenic Research 1:124-132 (1992); Ebert, K.M., et al., Biotechnology 12:699-702 (1994)). This limits the commercial utility of the transgenic production system because many commercially valuable proteins are enzymes, growth factors, or even toxins.

SUMMARY OF THE INVENTION

The invention provides a solution to the longstanding problem of inefficient or variable tissue-specific expression of heterologous genes in mammary gland tissue. Accordingly, the invention features transcription regulatory elements derived from a milk specific promoter, e.g., a mammalian lactoferrin gene promoter. An isolated nucleic acid within the invention containins a promoter region derived from the human lactoferrin gene operably linked to a heterologous sequence. A heterologous sequence is one that does not encode a naturally occurring lactoferrin polypeptide. The promoter region includes at least 20 nucleotides of the nucleotide sequence of SEQ ID NO:1. For example, the promoter region contains nucleotides 1-154 of SEQ ID NO:1 or 2.

Table 1: Human Lactoferrin promoter region

1 CT*GGATCC*TCAAGGAACAAGTAGACCTGGCCGCGGGGAGT
41 GGGGAGGGAAGGGGTGTCTATTGGGCAACAGGGCGGCAAA
81 GCCCTGAATAAAGGGGCGCAGGGCAGGCGCAAGTGCAGAG
121 CCTTCGTTTGCCAAGTCGC*CTCGAG*ACCGCAGACATGAAA
GCATGTCTCCGCGGAAAA (SEQ ID NO:1)

BamH1 restriction site GGATCC (nucleotides 5-8) and XhoI site (nucleotides 140-145) are italicized. These restriction sites may be altered, e.g., replaced with other restriction sites or with nucleotides that do not represent restriction enzyme recognition sites.

Table 2: Human Lactoferrin promoter region

1 CTXXXXXXTCAAGGAACAAGTAGACCTGGCCGCGGGGAGT
41 GGGGAGGGAAGGGGTGTCTATTGGGCAACAGGGCGGCAAA
81 GCCCTGAATAAAGGGGCGCAGGGCAGGCGCAAGTGCAGAG
121 CCTTCGTTTGCCAAGTCGCXXXXXXACCGCAGACATGAAA
GCATGTCTCCGCGGAAAA (SEQ ID NO:2)

Optionally, the lactoferrin-derived promoter regions described above are linked to nucleotides 1-1176 of nucleotide sequence of SEQ ID NO:16 (GENBANK™ accession no. S52659).

Table 6

1 cgaggatcat ggctcactgc caccttcatc tcccaggctc aaatggtcct cccactttag
61 cctcccaagt agctgggacc ataggcatac accaccatgc tgggctaatt tttgtatttt
121 ttgtagagat gggggtttcc ctatgaagcc caggctagtc ttgaactcct gggctcaagc
181 gatcctccca tcttggcctc ccaaagtgct gggattacag gcatgagcca ctgtgccctg
241 cctagttact cttgggctaa gttcacatcc atacacacag gatattcttt ctgaggcccc
301 caatgtgtcc cacaggcacc atgctgtatg tgacactccc ctagagatgg atgtttagtt
361 tgcttccaac tgattaatgg catgcagtgg tgcctggaaa catttgtacc tggggtgctg
421 tgtgtcatgg gaatgtattt acgagatgta ttcttagaag cagtattcta gcttttgaat
481 tttaaaatct gacatttatg gcgattgtta aaatgaggtt accatttcct attgaatact
541 atcaacacca aaaaagaaga aggaggagat ggagaaaaaa aagacaaaaa aaaaaaaagt
601 ggtagggcat cttagccata gggcatcttt ctcattggca aataagaaca tggaaccagc
661 cttgggtggt ggccattccc ctctgaggtc cctgtctgtt ttctgggagc tgtattgtgg
721 gtctcagcag ggcagggaga taccccatgg gcagcttgcc tgagactctg ggcagcctct
781 cttttctctg tcagctgtcc ctaggctgct gctgggggtg gtcgggtcat cttttcaact
841 ctcagctcac tgctgagcca aggtgaaagc aaacccacct gccctaactg gctcctaggc
901 accttcaagg tcatctgctg aagaagatag cagtctcaca ggtcaaggcg atcttcaagt
961 aaagaccctc tgctctgtgt cctgccctct agaaggcact gagaccagag ctgggacagg
1021 gctcaggggg ctgcgactcc tagggggcttg cagacctagt gggagagaaa gaacatcgca
1081 gcagccaggc agaaccagga caggtgaggt gcaggctggc tttcctctcg cagcgcggtg

1141 tggagtcctg tcctgcctca gggcttttcg gagcctggat cctcaaggaa caagtagacc
1201 tggccgcggg gagtggggag ggaaggggtg tctattgggc aacagggcgg ggcaaagccc
1261 tgaataaagg ggcgcagggc aggcgcaagt ggcagagcct tcgtttgcca agtcgcctcc
1321 agaccgcaga catgaaactt gtcttcctcg tcctgctgtt cctcggggcc ctcggtgagt
1381 gcaggtgcct gggggcgcga gccgcctgat gggcgtctcc tgcgccctgt ctgctaggcg
1441 ctttggtccc tgtgtccggt tggctgggcg cggggtctct gcgccccgcg gtcccagcgc
1501 ctacagccgg gaggcggccc ggacgcgggg ccagtctctt tcccacatgg ggaggaacag
1561 gagctgggct cctcaagccg gatcggggca cgcctagctc tgctcagagc ttctcaaaag
1621 gcctcccagg cccctgtccc tttgtgtccc gcctaaggat ttggtcccca ttgtattgtg
1681 acatgcgttt tacctgggag gaaagtgagg ctcagagagg gtgagcgact agctcaagga
1741 ccctagtcca gatcctagct cctgcgagga ctgtgagacc ccagcaagac cgagccttta
1801 tgagacttag tttcttcact taaagaaacg gcctaaccat gggtccacag ggttgtgagg
1861 aggagatggg gcattcgcac accttccgtg gcagagggtt gtggaggggt gcggtgctcc
1921 tgatggaacc ctgtgtcaga gggtttgaga gggaaatgtc agccaaacag aaggaaggag
1981 cagaaggaag gaaacaattg tcagttccat aaccaaagta atttctcggg tgctcagagg
2041 gcactcccca gcgctgcaca ttagtgacct aaatgcgtga gtgcgg (SEQ ID NO: 16)

By "isolated" is meant a nucleic acid molecule that is free of the genes which, in the naturally-occurring genome of the organism, flank the sequence of interest. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a procaryote or eucaryote; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence. The term excludes large segments of genomic DNA, e.g., such as those present in cosmid clones, which contain a given DNA sequence flanked by one or more other genes which naturally flank it in a naturally-occurring genome.

The lactoferrin-derived transcription regulatory sequences, are attached to a nominal promoter (e.g., the nominal lactoferrin promoter or a heterologous promoter) which in turn is operably linked to a sequence to be transcribed. The heterologous sequence to be transcribed is a polypeptide-encoding sequence or antisense sequence. When incorporated into a transgenic mammal such as a cow, the regulatory sequences of the invention operably linked to a polypeptide-encoding sequence direct expression of a polypeptide at a level of at

least 0.1 mg/ml in milk. Preferably, the sequences direct production of the trangene product at a level of 1-5 mg/ml in milk.

The regulatory sequences described herein may be used as a bi-directional promoter capable of exerting its function independently of its orientation in relation to the nucleic acid to be transcribed. A nucleic acid according to the invention is obtained by any technique in use in the art, for example by cloning, hybridization with the aid of an appropriate probe, by Polymerase Chain Reaction (PCR), or by chemical synthesis.

The nucleic acid of the invention includes an RNA stabilization sequence and/or a polyadenylation (poly A) sequence. Such stabilization or poly A sequences are preferably operatively linked to the heterologous nucleic acid sequence at the 3' end of the sequence to be transcribed. The heterologous nucleic acid to be transcribed is preferably insulin, calcitonin, serum albumin, a tetrameric antibody, an FAb fragment, a single chain antibody, a plasma protein, an industrial enzyme, silk, or a membrane receptor. The RNA stabilization sequence includes nucleotides 424-1058 of SEQ ID NO:3 or 4.

Table 3: 3' Region of Human Lactoferrin Gene

1	CAGGNTGGCCCAGTAAGGATTCCTGNGAATGAATTGAGTG
41	AATCTGCCAGGTGAACATGGATTGCAAACCGGGTTCACAT
81	TCCCCGGNAGAAGCTAGAGGNCCCACCCAATTTCTTGTGA
121	ACTTGAGAATGTGACAGTCGATTCAATCAGAGACAAGTGC
161	AGGGTGGTTGTGTCTCTCAGGCCAGAGCAGGGAAACACCC
201	TGGCTGGTGAGGGCTAGACTCTGGCTCCCTTGAACACCGT
241	AGTCGCTAGGAGTAGGGGAGTGGGAATATGAGTGTGGCAA
281	GCACTGACTCAGTGATGGGAGAAGGGCAGAGAAAACTCTT
321	AGTATTCTCTTTGATTTATTGGATTAAATAACTGGTTTAA
361	TGGAAGAAATCAGTTTCTGAATCTCTTGCTCTGTTGTGTC
401	CCACAGCCCTCCTGGAAGCCTGTGAATTCCTCAGGAAGTA
441	AAACCGAAGAAGATGGCCCAGCTCCCCAAGAAAGCCTCAG
481	CCATTCACTGCCCCCAGCTCTTCTCCCCAGGTGTGTTGGG
521	GCCTTGGCCTCCCCTGCTGAAGGTGGGGATTGCCCAT

561 CCATCTGCTTACAATTCCCTGCTGTCGTCTTAGCAAGAAG
601 TAAAATGAGAAATTTTGTTGATATTCTCTCCTTATAAAGT
641 GTCACTCATCTTTTCTAGAATTTTATACTGAAATCACATG
681 CCTGACAAAATACCTGTACAGTTGGACCTTCCCTTCCAAG
721 TTTTCAGGTCCAGCCCCTCCTCTTTCTTGCAGTCTTGGGT
761 ATGATGCCCAAGGGTCTGGAATTTAAGGCCAGGCCAAGCA
801 CCGGTTTTCCTAAGGGGATCTTGGTGGGTTATTCACATAG
841 CTGGCTCANTGCACGTGCATGTATGTGCCTGGGAATGTNT
881 GCCNTGTCCCCAAGGCAGGGCAGGGAAAGACCAAGGCCTT
921 GGGAAATTATTAACNGGAAANNTANGGGTTCCAANTNGCC
961 NCAATCNCNTTGCNNAAGTCCTAAATTTAACCAAGANCCT
1001 NGGGTTGGGGTTTAAAAAGGGGGACCTTTTAATTCCCNAA
1041 AGNTTCCCCTTAGGGGGG.....TGCGACAAGCCGC
CGAAAGTTCCTCGAAGCTAGCTTCAGACGTGTCTAGA

(SEQ ID NO: 3); bold type indicates nucleotides in exon 17 of human lactoferrin gene; "..." indicates a gap.

Table 4: Lactoferrin-derived RNA stabilization sequence

XXXXXXXAATTCCTCAGGAAGTA
AAACCGAAGAAGATGGCCCAGCTCCCCAAGAAAGCCTCAG
CCATTCACTGCCCCCAGCTCTTCTCCCCAGGTGTGTTGGG
GCCTTGGCCTCCCCTGCTGAAGGTGGGGATTGCCCAT
CCATCTGCTTACAATTCCCTGCTGTCGTCTTAGCAAGAAG
TAAAATGAGAAATTTTGTTGATATTCTCTCCTTATAAAGT
GTCACTCATCTTTTCTAGAATTTTATACTGAAATCACATG
CCTGACAAAATACCTGTACAGTTGGACCTTCCCTTCCAAG
TTTTCAGGTCCAGCCCCTCCTCTTTCTTGCAGTCTTGGGT
ATGATGCCCAAGGGTCTGGAATTTAAGGCCAGGCCAAGCA
CCGGTTTTCCTAAGGGGATCTTGGTGGGTTATTCACATAG
CTGGCTCANTGCACGTGCATGTATGTGCCTGGGAATGTNT
GCCNTGTCCCCAAGGCAGGGCAGGGAAAGACCAAGGCCTT

GGGAAATTATTAACNGGAAANNTANGGGTTCCAANTNGCC
NCAATCNCNTTGCNNAAGTCCTAAATTTAACCAAGANCCT
NGGGTTGGGGTTTAAAAAGGGGGACCTTTTAATTCCCNAA
AGNTTCCCCTTAGGGGGG (SEQ ID NO:4)

The stabilization sequence may optionally include the nucleotide sequence TGCGACAAGCCGCCGAAAGTTCCTCGAAGCTAGCTTCAGACGTGTCTAGA (SEQ ID NO:5).

Also within the invention is an isolated nucleic acid containing a lactoferrin-derived dominant control region (DCR) in the presence or absence of a lactoferrin-derived promoter sequence. A DCR is a nucleic acid sequence which directs consistent level, site of integration-independent, copy number-dependent expression of a nucleic acid operably linked thereto. For example, a DCR derived from genomic DNA located 5' or 3' to the transcription start site of lactoferrin directs transcription of a transgene product in mammary gland tissue of a transgenic mammal. Alternatively, the DCR confers inducibility of polypeptide-encoding sequence to which it is linked. Preferably, the DCR regulates tissue-specific transcription of a heterologous nucleic acid sequence; the regulation of transcription by is position independent relative to the location of the heterologous nucleic acid sequence. For example, the DCR is located 5' or 3' to the sequence to be transcribed. An increase in the level of transcription of a heterologous nucleic acid sequence under the control of a DCR is directly proportionate to the number of copies of the DCR.

A nucleic acid is a nucleotide polymer, e.g, a DNA or RNA. Preferably, the nucleic acid is a double-stranded DNA.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

Fig. 1 is a diagram of the human lactoferrin gene locus and a representation of overlapping BAC clones. The shaded box represents the lactoferrin coding sequence and the hatched box represents dominant control regions.

Figs. 2A is a diagram of of a human lactoferrin PAC clones, and Fig. 2B is a diagram of human lactoferrin PAC subclones. B = BamHI, R = EcoRI, Sp = SpHI, X = XbaI, Xh = XhoI.

Fig. 3 is a diagram of the construction strategy for a human lactoferrin expression cassette. B = BamHI, R = EcoRI, N = NotI, S = Sal ISp = SpHI, X = XbaI, Xh = XhoI

DETAILED DESCRIPTION

Human lactoferrin genomic DNA was cloned, and a milk specific expression cassette constructed utilizing human lactoferrin promoter sequences and other lactoferrin-derived enhancer and regulatory elements. Lactoferrin is found in concentrations of at least 2 mg/ml in human breast milk which makes it a minor component of milk (Masson, P.L. and Heremans, J.F. Comp. Biochem. Physiol. 39B:119-129 (1971)). The lactoferrin promoter is a moderate strength promoter when compared to the casein promoters which direct high level expression of casein (10-20 mg/ml). In addition, the human lactoferrin promoter is somewhat unique compared to lactoferrin promoters of other species which direct dramatically lower levels of lactoferrin in milk. The human lactoferrin promoter is an optimal promoter for directing expression of heterologous proteins in mammary gland tissue of transgenic animals.

The human lactoferrin locus (Fig. 1) was isolated from commercially available human bacterial artificial chromosome (BAC) human P1 artificial chromosome (PAC) libraries. Due to the unique nature of the BAC and PAC clones, the entire locus was covered in 2-5 individual clones. Each clone is capable of holding 75-150 kb of genomic DNA unlike cosmid vectors which can only hold 30-40 kb. The clones from the different libraries were characterized by restriction analysis and southern blotting to ensure that overlapping clones were isolated (Fig. 1). These overlapping clones were used to construct a milk specific expression cassette and to isolate the dominant control region for the locus.

The human lactoferrin gene along with 20-30 kb of surrounding flanking sequence was subcloned from one of the artificial chromosome vectors into a cosmid vector. The gene was engineered to delete the protein coding sequence and add unique cloning sites for the addition of heterologous protein coding sequences. The human lactoferrin promoter is used to direct expression of foreign proteins to the milk of transgenic non-human mammals. The promoter is attached to either genomic or cDNA protein coding sequences. The human

lactoferrin 3' flanking sequence or a 3' flanking sequence of any other gene is inserted into the expression cassette or vector to ensure stable mRNA expression and poly adenylation. For example, the 3' flanking sequence is derived from the 3' flanking region of actin, albumin, or butyrophilin.

The transcription unit of the transgene expression system of the invention contains DNA sequences encoding a transgene, any expression control sequences such as a promoter or enhancer, a polyadenylation element, and any other regulatory elements that may be used to modulate or increase expression, all of which are operably linked in order to allow expression of the transgenic polypeptide.

Preferably, the human lactoferrin promoter regulatory DNA is used to control expression of a transgene in a transcription unit, or a truncated fragment of this promoter which functions analogously may be used. The lactoferrin-derived regulatory sequence, e.g., promoter sequence or DCR is positioned 5' to a heterologous nucleic acid sequence, e.g., a transgene, in a transcription unit. Portions of the lactoferrin-derived promoter region are tested for their ability to allow tissue-specific and elevated expression of a transgene using assays known in the art, e.g., standard reporter gene assays using luciferase, beta-galactosidase, or expression of an antibiotic resistance gene as a detectable marker for transcription. All or part of one of the nucleotide sequences specified in a reference sequence, e.g., SEQ ID NO:1 or 2, its complementary strand or a variant thereof may be used in to direct transcription of a heterologous nucleic acid sequence such as a transgene in a transgenic mammal. A nucleic acid fragment is a portion of at least 20 continuous nucleotides identical to a portion of length equivalent to one of the reference nucleotide sequences or to its complement.

The invention includes sequences which hybridize under stringent conditions, with all or part of the sequence reported in a reference sequence and retains transcription regulatory function. For example, the nucleic acid may contain one or more sequence modifications in relation to a reference sequence. Such modifications may be obtained by mutation, deletion and/or addition of one or more nucleotides compared to the reference sequence. Modifications are introduced to alter the activity of the regulatory sequence, e.g., to improve promoter activity, to suppress a transcription inhibiting region, to make a constitutive promoter regulatable or vice versa. Modification are also made to introduce a restriction site facilitating

subsequent cloning steps, or to eliminate the sequences which are not essential to the transcriptional activity. Preferably, a modified sequence is at least 70% (more preferably at least 80%, more preferably at least 90%, more preferably at least 95%, more preferably at least 99%) identical to a reference sequence. The modifications do not substantially alter the transcription promoter function associated with the reference sequence (or a naturally-occurring lactoferrin promoter sequence). For example, modifications are engineered to avoid the site of initiation of translation.

Nucleotide and amino acid comparisons are carried out using the Lasergene software package (DNASTAR, Inc., Madison, WI). The MegAlign module used was the Clustal V method (Higgins et al., 1989, CABIOS 5(2):151-153). The parameter used were gap penalty 10, gap length penalty 10.

Alternatively, the nucleic acids described herein hybridize at high stringency to a strand of DNA having the reference sequence, or the complement thereof and have transcription regulatory activity. Hybridization is carried out using standard techniques, such as those described in Ausubel et al. (Current Protocols in Molecular Biology, John Wiley & Sons, 1989). "High stringency" refers to nucleic acid hybridization and wash conditions characterized by high temperature and low salt concentration, i.e, hybridization at 42 degrees C, and in 50% formamide; a first wash at 65 degrees C, 2X SSC, and 1% SDS; followed by a second wash at 65 degrees C and 0.2% x SSC, 0.190 SDS. Lower stringency conditions suitable for detecting DNA sequences having about 50% sequence identity to a reference gene or sequence are detected by, for example, hybridization at 42 degrees C in the absence of formamide; a first wash at 42 degrees C, in 6X SSC, and 1% SDS; and a second wash at 50 degrees C, in 6X SSC, and 1% SDS.

Techniques to evaluate whether a variant has a promoter activity or transcription regulatory activity are known in the art. For example, the sequence to be tested is inserted upstream of a reporter gene whose expression is detectable (e.g., β-galactosidase, catechol oxygenase, luciferase or a gene conferring resistance to an antibiotic). The promoter activity or transcription regulatory activity is at least 50% (more preferably 60%, more preferably 70%, more preferably 80%, more preferably 90%, more preferably 95%, more preferably 99%, and most preferably 100%) of that associated with the reference sequence (or a naturally-occurring lactoferrin promoter or DCR). A sequence may also be modified so that

the promoter activity or transcription regulatory activity is greater than that associated with the reference sequence (or a naturally-occurring lactoferrin promoter or DCR). For example, an increase in promoter activity is at least twice that of naturally-occurring lactoferrin sequence. In another example, an increase in transcriptional acitivity is directly proportionate to the number of copies of a given regulatory sequence, e.g., a DCR. Thus, a transcription unit or expression cassette may contain two or more copies of a regulatory sequence such as a DCR in tandem to increase production of a desired gene product.

The components of a transgene expression system are delivered to a cell on one or more vectors, which include, but not limited to, plasmids and viruses. One or more transcription units may be provided on a plasmid, where a lactoferrin-derived promoter region is used to control expression and is positioned 5' to a transgene

Example 1: Identification of Dominant Control Regions (DCR)

In addition to the construction of a milk specific expression cassette, the isolated genomic clones are used to screen for a dominant control region (DCR) necessary for position independent, copy number dependent expression. To screen for a DCR, the two most distal clones which contain an intact lactoferrin coding sequence are isolated from all bacterial sequences and microinjected into a mammalian transgenic model, e.g., mouse or rat embryos, to produce transgenic animals. Any clone containing a DCR will produce equivalent amounts of human lactoferrin in the milk of all transgenic lines tested. Once a DCR has been localized to a single BAC or PAC clone, the DCR can be further localized by deletional analysis and the production of additional transgenic animals. Once the DCR has been localized to a 5-10 kb region, this region can be connected to the lactoferrin promoter cassette to direct position independent, copy number dependent expression. Although this procedure is described for the human lactoferrin locus, the technique is applicable to any locus such as but not limited to casein or lactoglobulin loci.

The lactoferrin-derived regulatory sequences described herein are useful to direct expression of a transgene in mammary gland tissue of a transgenic non-human mammal. The mammary gland is used as a bioreactor to produce commercially valuable proteins. The methods described herein are used to clone the human lactoferrin gene and surrounding dominant control elements of the lactoferrin gene as well as casein and whey protein loci to obtain consistent tissue-specific expression of heterologous proteins in mammary glandtissue.

Example 2: Isolation of Genomic Human Lactoferrin Clones

A milk specific promoter construct containing lactoferrin-derived transcription regulatory sequences is used for the production of foreign proteins in the milk of transgenic non-human mammals. The human lactoferrin gene was cloned and regulatory sequences modified for use as a promoter. The strategy described below is useful for isolating a milk specific dominant control region from any milk gene locus.

Human BAC and PAC libraries were purchased from Genome Systems Inc., St Louis, MO and were pre-blotted on to filters for screening. The filters were probed with oligonucleotides complimentary to the first and last exons of the lactoferrin gene. Reference sequences were obtained through the GENBANK™ system. All clones isolated were characterized by restriction analysis and southern blotting to determine regions of overlap.

Table 4: Oligonucleotides Used to Screen Human PAC Genomic Library

3'mRNA primers:

HLAC5	5'-GGAAGCCTGTGAATTCCTCAGGAA-3' (SEQ ID NO:6)
HLAC6	5'-GCAGGGAATTGTAAGCAGATGGAT-'3 (SEQ ID NO:7)

Promoter primers:

HLAC12	5'-CCTTGAGGATCCAGGCTCCGAA-3' (SEQ ID NO:8)
HLAC13	5'-GAAGATAGCAGTCTCACAGGTCAA-3' (SEQ ID NO:9)

Genomic clones containing the human lactoferrin gene were isolated using DOWN TO EARTH™ human PAC DNA pools purchased from Genome Systems, Inc. (St. Louis, MO). The human PAC DNA are arrayed in 20 microtiter dishes which can be screened using e consecutive rounds of PCR to identify individual clones of interest. The PAC library was constructed by ligating a partial Sau3A I digest of human DNA into the vector pAd10SacBII. The pAd10SacBII vector is a low P1 phage derived artificial chromosome vector capable of replication inserts of average size of 120 kb in the appropriate bacterial host. The vector is deisgned with T7 and SP6 promoters to enable sequencing of isolated clones and for chromosome walking in order to isolate entire gene loci or gene families.

In order to isolate the human lactoferrin gene, oligonucleotides were designed which were complimentary to the promoter region (sequence derived from GENBANK™ Accession #S52659) and the 3' end of the human lactoferrin mRNA (sequence derived from GENBANK™ Accession # X53961) for use in a polymerase chain reaction (see Table 4). The PCR primers were tested utilizing human genomic DNA and found to generate PCR fragments of the predicted size. The primers HLAC5 and HLAC6 were then used to screen the human PAC DNA pools and two positive clones were identified. The two clones were localized to wells 94K13 and 169a20 and ordered from Genome Systems, Inc. The bacterial clones were grown under kanamycin selection an amplified using IPTG for large scale preparation according to the manufacturer's protocol. To ensure that the clones contained the entire human lactoferrin gene, the two clones were then screened by PCR using the HLAC12 and HLAC13 primers. Both clones were found to contain the full length human lactoferrin gene and were then used for restriction mapping and subcloning of the gene fragments for construction of a mammary gland specific expression cassette.

Example 3: Construction of a Mammary Gland Specific Expression Cassette

To construct a mammary gland specific expression cassette, the promoter and 3' flanking regions of the human lactoferrin gene were subcloned and unique restriction enzyme sites added to allow for the addition of heterologous coding sequences and excision from the vector backbone. A schematic representation of the two human lactoferrin clones is shown in Fig. 2A (not drawn to scale). Each clone contained an insert of approximately 120 kb. The human lactoferrin gene is approximately 24.5Kb in length and is divided into 17 exons (Kim et al., Mol. Cells 8(6):663-8 (1998). As shown in Fig. 2B, the human lactoferrin gene was subcloned as five distinct fragments into the vectors pUC19 (New England BioLabs, Beverly, MA) or Sc1. The cosmid Sc1 was derived from the vector Supercos (Stratagene, La Jolla, CA) and has a multiple cloning site (SalI-BamHI-XhoI-NotI) added between the two EcoRI sites. The subclones were then used to reassemble a mammary gland specific expression cassette of the human lactoferrin gene.

The promoter region was reconstructed as a SalI to XhoI fragment using the subclones HL3 and HL10 (Fig. 3). A unique XhoI restriction site was added before the ATG initiation codon using polymerase chain reaction mutagenesis and the oligonucleotides HL14 and HL14 (Table 5). The 500 bp PCR fragment amplified from the vector HL3 was subcloned into PvuII

digested pUC19 to form the vector HL12. The plasmid HL12 was then digested with BamHI and XhoI to excise the human lactoferrin fragment which was ligated into BamHI/XhoI digested Sc1 to form the vector HL14. HL14 was digested with BamHI, treated with calf intestinal alkaline phosphatase, and the 3.2 kb fragment from HL10 inserted. The orientation of the 3.2 kb insert was determined by restriction analysis and confirmed by DNA sequencing. The final vector was designated HL15 and contains approximately 3 kb of promoter sequence which can be excised as a SalI to XhoI fragment.

Table 5: Oligonucleotides Used to Add an XhoI site Upstream of the Initiation Codon

HLAC14 5'-CCTTCAAGGTCGACTGCTGAAGAAGAT-3' (SEQ ID NO:10)

HLAC17 5'-CATGTCTGCGGTCTCGAGGCGACTTGGCAA-'3 (SEQ ID NO:1)

HLLINK3 5-CTAGATAAGCCGACTCCAGCAGTAACGTCGACGCGGCCGCA-3' (SEQ ID NO:12)

HLLINK4 5'-AGCTTGCGGCCGCGTCGACGTTACTGCTGGAGTCGGCTTAT-3' (SEQ ID NO:13)

The 3' flanking region of the gene was subcloned as single BamHI fragment of over 20 kb in length which was designated HL11 (Fig. 3). Restriction analysis of the vector HL11 revealed the presence of several XhoI sites which were removed before reconstruction of the 3' flanking region. To remove the XhoI sites, the 3' end was further subcloned by digestion with EcoRI or XbaI into the vector pUC19. Two overlapping clones designated HL16 and HL24 were found to contain the stop codon and immediate 3' region of the gene. In order to add a unique 3' restriction site, the plasmid HL16 was digested with XbaI which leaves the 5'fragment attached to the vector backbone, gel purified, and ligated with a synthetic linker (Table 5, oligonucleotides HLLINK3 and HLLINK4). The correct orientation of the linker was determined by restriction analysis and the new plasmid designated HL26. The plasmid HL26 was then digested EcoRI and ligated with the synthetic linker:

5'-AATTGCTCGAGC-3' (SEQ ID NO:14)

5'-CGAGCTCGTTAA-3' (SEQ ID NO:15)

The addition of the linker converts the EcoRI site to an XhoI site and forms the plasmid HL27. To complete the 3' flanking region, HL27 was digested with XbaI and SalI and ligated with the 7 kb XbaI/XhoI fragment from HL24. The final construct was designated HL28 and could be excised as an XhoI to NotI fragment approximately 7.2 kb in length. The XhoI/NotI fragment from HL28 was then ligated into XhoI/NotI digested HL15 to form the final vector HL29 (Fig. 3).

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid, comprising a promoter region derived from the human lactoferrin gene operably linked to a heterologous sequence, wherein said promoter region comprises nucleotides 1-154 of the nucleotide sequence of SEQ ID NO:2.

2. The nucleic acid of claim 1, wherein said nucleic acid further comprises nucleotide 1-1176 of the nucleotide sequence of SEQ ID NO:16.

3. An isolated nucleic acid, comprising a promoter region derived from the human lactoferrin gene operably linked to a heterologous sequence, wherein said promoter region comprises nucleotides 1-154 of the nucleotide sequence of SEQ ID NO:1.

4. The nucleic acid of claim 2, wherein said nucleic acid further comprises nucleotides 1-1176 of the nucleotide sequence of SEQ ID NO:16.

5. The nucleic acid of claim 1, wherein said heterologous sequence encodes a polypeptide.

6. The nucleic acid of claim 2, wherein said heterologous sequence does not encode a naturally occurring lactoferrin polypeptide.

7. The nucleic acid of claim 1, wherein said nucleic acid further comprises an RNA stabilization sequence.

8. The nucleic acid of claim 7, wherein said RNA stabilization sequence comprises nucleotides 424-1058 of the nucleotide sequence of SEQ ID NO:3.

9. The nucleic acid of claim 7, wherein said RNA stabilization sequence comprises the nucleotide sequence of SEQ ID NO:4.

10. The nucleic acid of claim 8, wherein said RNA stabilization sequence further comprises the nucleotide sequence of SEQ ID NO:5.

11. The nucleic acid of claim 9, wherein said RNA stabilization sequence further comprises the nucleotide sequence of SEQ ID NO:5.

12. The nucleic acid of claim 1, wherein said nucleic acid further comprising a polyadenylation sequence.

13. The nucleic acid of claim 1, wherein said heterologous sequence encodes be insulin, calcitonin, serum albumin, a tetramric antibody, an FAb fragment, a single chain antibody, a plasma protein, an industrial enzyme, silk, or a membrane receptor.

14. An isolated nucleic acid comprising a lactoferrin-derived promoter sequence and a dominant control region (DCR).

15. The nucleic acid of claim 8, wherein said DCR regulates tissue-specific transcription of a heterologous nucleic acid sequence, wherein regulation of transcription by said DCR is position independent relative to the location of said heterologous nucleic acid sequence.

16. The nucleic acid of claim 8, wherein said DCR regulates transcription of a heterologous nucleic acid sequence, wherein an increase in the level of transcription of said heterologous nucleic acid sequence is directly proportionate to the number of copies of said DCR.

17. A transgenic non-human mammal comprising the isolated nucleic acid of claim 1.

~~Figure 1~~. Fig. 1

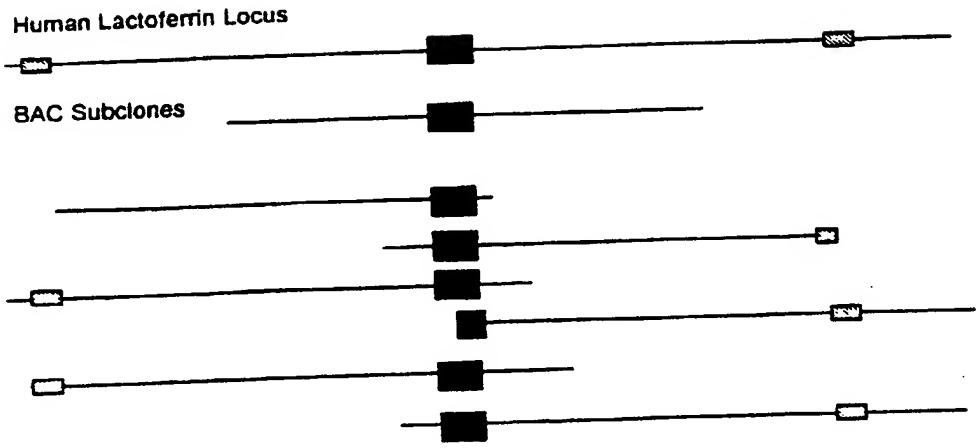


Fig. 2A

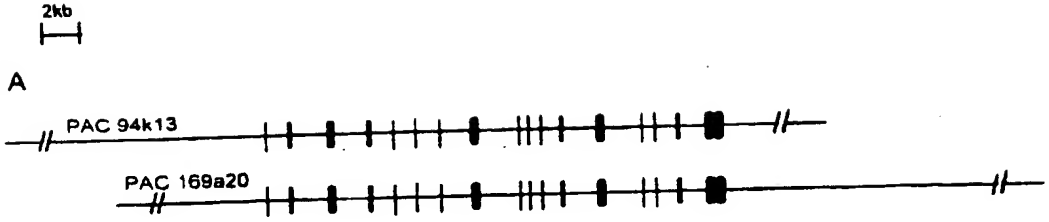

2kb
PAC 94k13
PAC 169a20

Fig. 2B

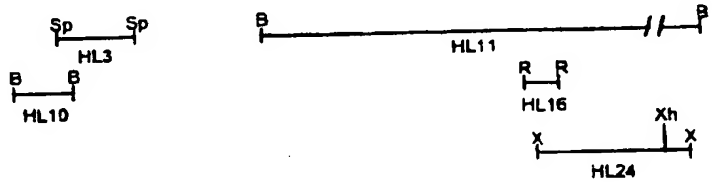
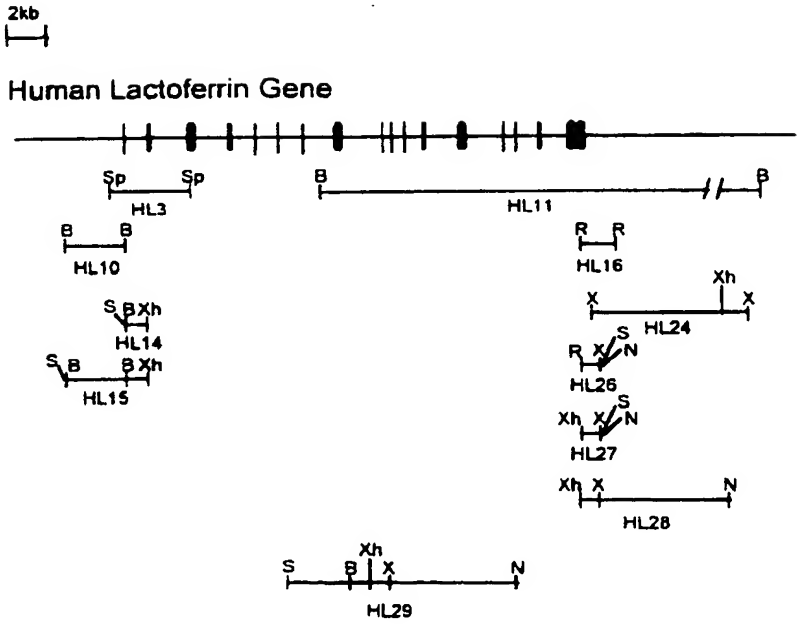

Sp
Sp
HL3
B
B
HL10
B
B
HL11
R
R
HL16
Xh
X
X
HL24

Fig 3


2kb
Human Lactoferrin Gene
Sp
Sp
HL3
B
B
HL11
B
B
HL10
R
R
HL16
S
B
Xh
HL14
Xh
X
X
HL24
S
R
X
N
HL26
S
B
B
Xh
HL15
S
Xh
X
N
HL27
Xh
X
N
HL28
Xh
S
B
X
N
HL29

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01662

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/00, 15/63, 15/09

US CL :526/25.1; 435/320.1, 455; 800/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 526/25.1; 435/320.1, 455; 800/14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, WEST, NPL

search terms: lactoferrin promoter, dominant control region, regulatory, transcription

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WANG et al. Characterization and functional analysis of the porcine lactoferrin gene promoter. Gene. 1998, Vol. 215, pages 203-212, see entire document.	14

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 13 JUNE 2000	Date of mailing of the international search report 05 JUL 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JILL D. MARTIN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01662

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-13 and 15-17
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

these claims recite SEQ ID NOS. which could not be searched because no nucleotide sequence listing and computer readable form have been furnished to the Office.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.